2018. *Proc Ling Soc Amer* 3. 24:1-9. https://doi.org/10.3765/plsa.v3i1.4311

Investigating a possible "musician advantage" for speech-in-speech perception: The role of f0 separation

Michelle D. Cohn[*]

Abstract. Does listeners' musical experience improve their ability to perceive speech-in-speech? In the present experiment, musicians and nonmusicians heard two sentences played simultaneously: a target and a masker sentence that varied in terms of fundamental frequency (f0) separation. Results reveal that accuracy in identifying the target sentence was highest for younger musicians (relative to younger nonmusicians). No such difference was observed between older musicians and nonmusicians. These results provide support for musicians' purported advantage for speech-in-speech – but the advantage is limited by listener age. This work is relevant to our understanding of cross-domain transfer of nonlinguistic experience on speech perception.

Keywords. speech-in-speech perception, fundamental frequency, cross-domain plasticity

1. Introduction. As listeners, we rarely experience ideal listening conditions; often we must contend with one or more competing speakers (e.g., in a busy café) to hear the target talker. Our ability to tease apart these competing voices, however, is not trivial. And listeners might vary in the strategies involved in successful speech-in-speech perception, such as in leveraging certain acoustic cues. Furthermore, with age, filtering out this noise (i.e., the competing signal) becomes an even more difficult task. An understanding of the mechanisms underlying successful speech perception amidst background babble – and sources of individual variation in this ability – are important in addressing this common concern for listeners and informing our models of speech in noise perception (e.g., Anderson et al. 2013).

A number of groups have shown that introducing acoustic differences between the target and competing voice(s) improves intelligibility, including spatial separation (Hawley, Litovsky & Colburn 1999), timing (Carhart, Tillman & Greetis 1969), amplitude (Brungart 2001), and fundamental frequency (f0) (Summerfield & Assmann 1991; Summers & Leek 1998). Furthermore, listeners' abilities to leverage these acoustic cues has shown to vary according to their linguistic backgrounds. For example, speech perception in multitalker babble is more difficult for non-native versus native speakers (e.g., Mayo, Florentine & Buus 1997) and for accents listeners have less experience with (e.g., for French speakers listening to French- vs. British-accented English in babble (Pinet & Iverson 2010)). Many listeners also have another type of experience that may impact their ability to perceive speech in the presence of background speakers: musical training.

Musicians have specialized auditory training involving fine-grained acoustic distinctions of musical sounds (e.g., pitch, amplitude, timing, etc.). Whether this experience can *transfer* to speech perception, however, is an unresolved question – some studies show musicians' enhanced speech-in-speech perception relative to nonmusicians (Parbery-Clark et al. 2009; Strait et al.

[*] Thank you to Dr. Georgia Zellou and Dr. Santiago Barreda for their guidance with the experiment and helpful comments on the manuscript and to Steven Ilagan, our undergraduate research assistant, for his help running the subjects. Author: Michelle D. Cohn, University of California, Davis (mdcohn@ucdavis.edu)

2013; Vasuki et al. 2016; Başkent & Gaudrain 2016; Zendel et al. 2017; Meha-Bettison et al. 2018), while others report no significant difference between these groups (Ruggles, Freyman & Oxenham 2014; Boebinger et al. 2015; Madsen, Whiteford & Oxenham 2017) or an enhancement limited by musicians' age (e.g. only for musicians age ≥40 in Zendel & Alain 2012). Yet, the majority of these studies did not control for f0 separation and fluctuation – two acoustic cues that lead to increased intelligibility in perceiving competing speakers (e.g., Summerfield & Assmann 1991; Patel, Xu & Wang 2010) or other speaker-related cues (e.g., using different talkers for target and masker(s) in Boebinger et al. 2015; Madsen, Whiteford & Oxenham 2017).

2. Present Experiment. To address this gap in the literature (i.e., the need to control for acoustic characteristics of the target and masker(s)), the present experiment investigates the role of f0 separation for speech-in-speech perception across the groups, controlling for the acoustic cues of f0 fluctuation, spatial separation, amplitude, as well as speaker-related cues by using the same talker for the target and masker. Our focus on f0 separation is based on empirical work showing musicians' enhanced encoding of f0 both in pure tones (Kishon-Rabin et al. 2011), but also in language (e.g., detecting weakly incongruous prosodic contours in Schön, Magne & Besson 2004). We hypothesize that musicians (relative to nonmusicians) will better leverage small f0 differences between competing talkers for improved speech-in-speech perception. Evidence for a transfer of skills developed in musical training, such as pitch perception, is supported by longitudinal studies that show increased fidelity in brainstem encoding of periodicity cues in speech sounds following with musical training (e.g., Kraus et al. 2014; Tierney, Krizman & Kraus 2015) and theoretical models of plasticity, such as the OPERA Hypothesis (Patel 2011). On the other hand, it is also possible that both musicians and nonmusicians will perform similarly on the task. That is, it is possible that *all* listeners will be equally sensitive to f0 manipulations, given the importance of f0 for intonational contours and as a cue for speaker gender.

2.1 SUBJECTS. Musicians (n=41) and nonmusicians (n=41) were all native English speakers with no prior experience with a tonal language and who reported normal hearing. Musicians had at least 10 years of musical training ($\bar{x}$=23.26 yrs, sd=14.59), while nonmusicians had minimal (<1 year) to no musical training. Additionally, subjects were matched in age: musicians ($\bar{x}$=39.73 yrs, sd=16.70, range=18-69) and nonmusicians ($\bar{x}$=38.22 yrs, sd=16.55, range=18-72). A t-test revealed no significant difference in age across the groups [$t(80) = 0.412$, $p = 0.682$].

2.2 STIMULI. Both target and masker sentences were selected from a single male speaker from the Coordinate Response Measure (CRM) database (Bolia et al. 2000). CRM sentences all have the same form: "Ready <call sign> go to <color> <number> now". Target sentences, indicated by the call sign "baron" (e.g., "Ready baron go to blue one now"), were monotonized at six f0 levels (relative to 100 Hz in 0, 0.156, 0.306, 1, 2, & 3 semitone increases) in Praat (Boersma & Weenink 2018). Masker sentences, containing different call signs (e.g., "Ready eagle..."), were monotonized at 100 Hz. Following a pseudorandom selection (with the restriction that the color/number had to differ between the target and masker), target and masker sentences were combined in R for a total of 96 tokens (16 targets x 6 f0 levels).

2.3 PROCEDURE. In a soundbooth at the UC Davis Phonetics Laboratory, subjects began with single sentence identification (12 trials) in which they clicked the color/number combination from the "baron" sentence using the onscreen grid of responses (see Figure 1). Following single sentence identification, subjects continued on to the experimental trials consisting of a target and masker sentence presented simultaneously (e.g., Target: "Ready baron go to blue one now" +

Masker: “Ready hopper go to green three now”). In the 192 experimental trials (8 blocks x 24 trials), the subjects’ task was again to select the color/number combination from the target “baron” sentence.

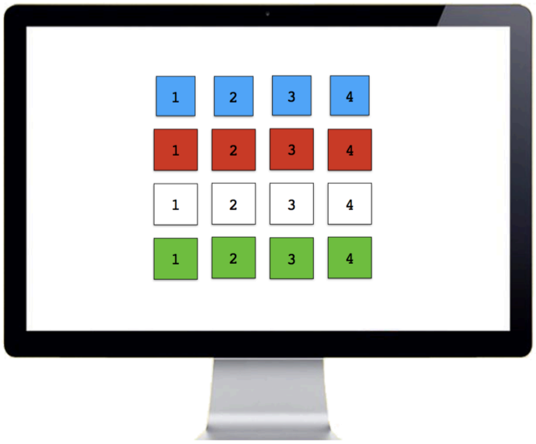

Figure 1: Example response screen

2.4 ANALYSIS. Accuracy in identifying both the color/number from the target was coded as 1 if both color & number identified, and 0 if not. Results were analyzed using a mixed effects logistic regression model with the lme4 R package (Bates, Bolker & Walker 2015). Fixed effects included F0 separation as a continuous variable, Group, and Age – with all possible two-way interactions (Group*Age, Group*F0, Age*F0) – and with a by-Subject random intercept and by-Subject random slope for F0 separation. Post-hoc analyses were conducted to test specific predictions of listener age (<40 vs. ≥40 per Zendel & Alain 2012) using two separate mixed effects models: one for younger listeners (<40) and one for older listeners (≥40). The covariates and random effects structure was identical to the main model, with the exception of Age (which was excluded).

3. Results. All subjects completed the single sentence portion of the study with 90% or greater (mean accuracy = 97.8%, sd = 0.15). Results from the mixed effects model (see Table 1) show that increasing F0 separation between the target & masker (p<0.001) and decreasing listener Age (p<0.001) significantly improve the identification of the color/number combination for the target sentence. No main effect of Group was observed (p=0.133).

	Probability of identifying target			
	Coef	*std. Error*	*Wald chisq*	*p*
F0	0.25	0.03	74.80	<0.001
Group$_{\text{MUS}}$	0.04	0.03	2.25	0.133
Age	-0.13	0.03	18.41	<0.001
F0*Group$_{\text{MUS}}$	0.01	0.03	0.15	0.698
F0*Age	-0.08	0.03	7.20	0.007
Group$_{\text{MUS}}$*Age	-0.05	0.02	4.23	0.040
Observations	15744			
Family	binomial (logit)			

Table 1: Mixed effects model output

Additionally, we see that the improvement observed based on F0 separation did not statistically differ between the Groups (Group*F0, N.S.) (see Figure 2).

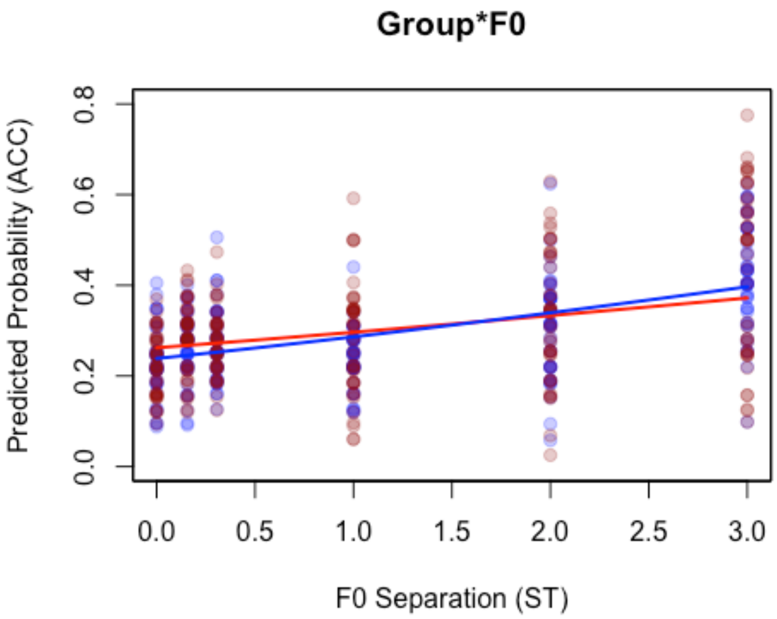

Figure 2: Interaction of Group and F0 separation in semitones. Mean accuracy for each subject shown as dots, with logistic functions plotted from the model output. (Musicians = red, Nonmusicians = blue).

On the other hand, we observed a significant interaction (p<0.01) between listener Age and F0 separation such that older listeners showed a weaker improvement in accuracy based on increasing F0 separation.

The model also showed a significant interaction between Group and Age (p<0.05), revealing that accuracy in identifying the target sentence was highest for younger musicians relative to younger nonmusicians (see Figure 3), but that musicians also had a steeper age-related decline.

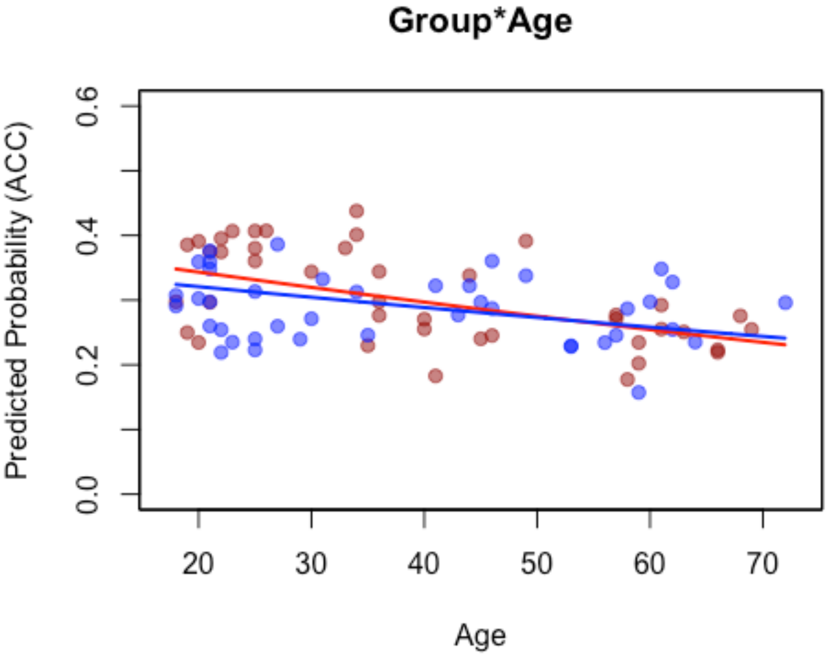

Figure 3: Interaction between Group and Age. Mean accuracy for each subject shown as dots, with logistic functions plotted from the model output. (Musicians = red, Nonmusicians = blue).

Post-hoc analyses testing the relationship between age category (<40, ≥40) and Group per Zendel & Alain (2012) revealed a significant difference between younger musicians and nonmusicians (p<0.001), but no difference between older musicians and nonmusicians (p=0.103) (see Table 2). For both subgroups (<40, ≥40), the interaction between F0*Group was not significant.

	Probability of identifying target							
	< 40				≥40			
	Coef	*std. Error*	*Wald Chisq*	*p*	*Coef*	*std. Error*	*Wald Chisq*	*p*
Group$_{\text{MUS}}$	0.13	0.04	10.85	<0.001	-0.07	0.04	2.65	0.103
F0	0.36	0.04	92.61	<0.001	0.16	0.05	11.30	<0.001
F0*Group$_{\text{MUS}}$	0.07	0.04	3.35	0.067	-0.07	0.05	1.87	0.171
Observations	8448 (n=44)				7296 (n=38)			
Family	binomial (logit)				binomial (logit)			

Table 2: Post-hoc mixed effects model outputs by group

4. Discussion. The purpose of this research was to investigate whether musicians are better at leveraging fundamental frequency (f0) separation for improved speech-in-speech perception. While linguistic experience is a strong predictor of successful speech-in-speech perception (e.g., Mayo, Florentine & Buus 1997), we tested whether nonlinguistic experience – in particular, musical training – would benefit listeners in speech-in-speech perception and whether any such benefit was moderated by listener age.

Our findings suggest that musical training may help listeners cue into f0 separation for successful speech-in-speech perception, but its effects are limited by age. While the main effect of Group was not significant, we observed a significant interaction between Group and Age ($p<0.05$), with a "musician's advantage" for younger listeners on the basis of f0 separation. This was confirmed by post-hoc analyses (see Table 2) in which younger musicians (< 40) significantly outperformed younger nonmusicians ($p<0.001$), but no such difference was observed for older musicians (≥ 40) and older nonmusicians ($p=0.103$). On the one hand, these results are consistent with theoretical frameworks of cross-domain auditory plasticity, wherein extensive training in one domain (e.g., music) may *transfer* to speech perception (e.g., Patel 2011) and empirical work showing that musical training "interventions" can lead to significantly more robust f0 encoding in the brainstem (Kraus et al. 2014; Tierney, Krizman & Kraus 2015) .

On the other hand, our results suggest that musicians may be *losing* their f0-based advantage with age – contra what was found in Zendel & Alain (2012). However, it is possible to reconcile these seemingly contradictory findings. The "musician advantage" for speech-in-speech perception observed with increasing age (e.g., Zendel & Alain 2012) may be due to other acoustic cues in the stimuli (e.g., amplitude, f0 fluctuation, spatial separation) to help separate the target from masker talker(s). While the current study sheds light on listeners' reliance on f0 separation for speech-in-speech perception, additional study is needed to tease apart the contributions of other acoustic cues across the lifespan and examine their possible interaction with listeners' musical background.

Overall, this study serves as a step toward (1) evaluating the purported “musicians’ advantage” for speech-in-speech in general, but also based on listeners’ age, and (2) more precisely testing the role of specific phonetic cues in speech-in-speech than what has been previously observed in the literature. Still, further work is needed to clarify the role of other cues and experience on the dynamic and difficult process of perceiving speech with background speaker(s).

5. Conclusion. The present experiment demonstrates that musicians and nonmusicians differ in their ability to perceive speech-in-speech on the basis of f0 separation, with their accuracy heavily modulated by their age. While younger musicians show an advantage compared to younger nonmusicians (under age 40), no difference is observed for musicians and nonmusicians over age 40. This work supports models of cross-domain plasticity in showing the influence of musical training on speech perception, but also suggests that the “musician’s advantage” based on f0 may decline over the lifespan.

References

Anderson, Samira, Travis White-Schwoch, Alexandra Parbery-Clark & Nina Kraus. 2013. A dynamic auditory-cognitive system supports speech-in-noise perception in older adults. *Hearing research* 300. 18. https:/doi.org/10.1016/j.heares.2013.03.006.

Başkent, Deniz & Etienne Gaudrain. 2016. Musician advantage for speech-on-speech perception. *The Journal of the Acoustical Society of America* 139(3). EL51-EL56. https:/doi.org/10.1121/1.4942628.

Bates, Douglas, Ben Bolker & Steve Walker. 2015. *Fitting linear mixed-effects models using lme4*. Journal of Statistical Software 67(1). 1–48. https://doi.org/10.18637/jss.v067.i01.

Boebinger, Dana, Samuel Evans, Stuart Rosen, César F. Lima, Tom Manly & Sophie K. Scott. 2015. Musicians and non-musicians are equally adept at perceiving masked speech. *The Journal of the Acoustical Society of America* 137(1). 378–387. https://doi.org/10.1121/1.4904537.

Boersma, Paul & David Weenink. 2018. *Praat: doing phonetics by computer.* http://www.praat.org/.

Bolia, Robert S., W. Todd Nelson, Mark A. Ericson & Brian D. Simpson. 2000. A speech corpus for multitalker communications research. *The Journal of the Acoustical Society of America* 107(2). 1065–1066. https://doi.org/10.1121/1.428288.

Brungart, Douglas S. 2001. Informational and energetic masking effects in the perception of two simultaneous talkers. *The Journal of the Acoustical Society of America* 109(3). 1101–1109. https://doi.org/10.1121/1.1345696.

Carhart, Raymond, Tom W. Tillman & Elizabeth S. Greetis. 1969. Release from Multiple Maskers: Effects of Interaural Time Disparities. *The Journal of the Acoustical Society of America* 45(2). 411–418. https://doi.org/10.1121/1.1911389.

Hawley, Monica L., Ruth Y. Litovsky & H. Steven Colburn. 1999. Speech intelligibility and localization in a multi-source environment. *The Journal of the Acoustical Society of America* 105(6). 3436–3448. https://doi.org/10.1121/1.424670.

Kishon-Rabin, L., O. Amir, Y. Vexler & Y. Zaltz. 2011. Pitch Discrimination: Are Professional Musicians Better than Non-Musicians? *Journal of Basic and Clinical Physiology and Pharmacology* 12(2). 125–144. https://doi.org/10.1515/JBCPP.2001.12.2.125.

Kraus, Nina, Jessica Slater, Elaine C. Thompson, Jane Hornickel, Dana L. Strait, Trent Nicol & Travis White-Schwoch. 2014. Music enrichment programs improve the neural encoding of speech in at-risk children. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 34(36). 11913–11918. https://doi.org/10.1523/JNEUROSCI.1881-14.2014.

Madsen, Sara M. K., Kelly L. Whiteford & Andrew J. Oxenham. 2017. Musicians do not benefit from differences in fundamental frequency when listening to speech in competing speech backgrounds. *Scientific Reports* 7(1). 12624. https://doi.org/10.1038/s41598-017-12937-9.

Mayo, Lynn Hansberry, Mary Florentine & Søren Buus. 1997. Age of Second-Language Acquisition and Perception of Speech in Noise. *Journal of Speech, Language, and Hearing Research* 40(3). 686–693. https://doi.org/10.1044/jslhr.4003.686.

Meha-Bettison, Kiriana, Mridula Sharma, Ronny K. Ibrahim & Pragati Rao Mandikal Vasuki. 2018. Enhanced speech perception in noise and cortical auditory evoked potentials in professional musicians. *International Journal of Audiology* 57(1). 40–52. https://doi.org/10.1080/14992027.2017.1380850.

Parbery-Clark, Alexandra, Erika Skoe, Carrie Lam & Nina Kraus. 2009. Musician Enhancement for Speech-In-Noise. *Ear and Hearing* 30(6). 653. https://doi.org/10.1097/AUD.0b013e3181b412e9.

Patel, Aniruddh D. 2011. Why would Musical Training Benefit the Neural Encoding of Speech? The OPERA Hypothesis. *Frontiers in Psychology* 2. https://doi.org/10.3389/fpsyg.2011.00142.

Patel, Aniruddh D., Yi Xu & Bei Wang. 2010. The role of F0 variation in the intelligibility of Mandarin sentences. *SP-2010*. paper 890.

Pinet, Melanie & Paul Iverson. 2010. Talker-listener accent interactions in speech-in-noise recognition: Effects of prosodic manipulation as a function of language experience. *The Journal of the Acoustical Society of America* 128(3). 1357–1365. https://doi/org/10.1121/1.3466857.

Ruggles, Dorea R., Richard L. Freyman & Andrew J. Oxenham. 2014. Influence of Musical Training on Understanding Voiced and Whispered Speech in Noise. *PLoS ONE* 9(1). e86980. https:/doi.org/10.1371/journal.pone.0086980.

Schön, Daniele, Cyrille Magne & Mireille Besson. 2004. The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology* 41(3). 341–349. https://doi.org/10.1111/1469-8986.00172.x.

Strait, Dana L., Alexandra Parbery-Clark, Samantha O'Connell & Nina Kraus. 2013. Biological impact of preschool music classes on processing speech in noise. *Developmental Cognitive Neuroscience* 6. 51–60. https://doi.org/10.1016/j.dcn.2013.06.003.

Summerfield, Quentin & Peter F. Assmann. 1991. Perception of concurrent vowels: Effects of harmonic misalignment and pitch-period asynchrony. *The Journal of the Acoustical Society of America* 89(3). https://doi.org/10.1121/1.400659.

Summers, Van & Marjorie R. Leek. 1998. F0 Processing and the Seperation of Competing Speech Signals by Listeners With Normal Hearing and With Hearing Loss. *Journal of Speech, Language, and Hearing Research* 41(6). 1294–1306. https://doi.org/10.1044/jslhr.4106.1294.

Tierney, Adam T., Jennifer Krizman & Nina Kraus. 2015. Music training alters the course of adolescent auditory development. *Proceedings of the National Academy of Sciences* 112(32). 10062–10067. https://doi.org/10.1073/pnas.1505114112.

Vasuki, Pragati Rao Mandikal, Mridula Sharma, Katherine Demuth & Joanne Arciuli. 2016. Musicians' edge: A comparison of auditory processing, cognitive abilities and statistical learning. *Hearing Research* 342. 112–123. https://doi.org/10.1016/j.heares.2016.10.008.

Zendel, Benjamin Rich & Claude Alain. 2012. Musicians experience less age-related decline in central auditory processing. *Psychology and Aging* 27(2). 410–417. https://doi.org/10.1037/a0024816.

Zendel, Benjamin Rich, Greg L. West, Sylvie Belleville & Isabelle Peretz. 2017. Music training improves the ability to understand speech-in-noise in older adults. *bioRxiv*. 196030. https://doi.org/10.1101/196030.